

ABSTRACT OF THE DISCLOSURE

5 A method for generating a language component vocabulary VC for a speech recognition system having a language vocabulary V of a plurality of word forms is disclosed. The method includes: partitioning the language vocabulary V into subsets of word forms based on frequencies of occurrence of the respective word forms; and in at least one of the subsets, splitting word forms having frequencies less than a threshold to thereby generate word form components. Also disclosed is a method for use in speech recognition including: splitting an acoustic vocabulary comprising baseforms into baseform components and storing the baseform components; and, performing sound to spelling mapping on the baseform components so as to generate a baseform components to word parts table for use in subsequent decoding of speech. A method for decoding a speech utterance using language model components and acoustic components, includes the steps of: generating from the utterance a stack of baseform component paths; concatenating baseform components in a path to generate concatenated baseforms, when the concatenated baseform components correspond to a baseform found in an acoustic vocabulary; mapping the concatenated baseforms into words; computing language model (LM) scores associated with the words using a language model, and performing further decoding of the utterance based thereupon.